8/2/2022. LECTURE 1: NOTES MADE ON THE VISUALISER.

INTRODUCTION.

Consider a problem where we wish to make INFERENCE about a parameter θ given observations, or DATA, x

CLASSICAL SETTING. (LIKELIHOOD BASED INFERENCE)

Data is treated as if it is random even after it has been observed.

Parameter is viewed as fixed unknown constant. Consequently, no probability distribution can be attached to the parameter.

BAYESIAN APPROACH

PARAMETERS having not yet been observed, are treated as random variables. Therefore, they possess a probability distribution.

Data, having been observed, is treated as fixed.

Example.

Consider n independent Bernoulli trials in which we observe x, the number of times an event occurs. We are interested in making inferences about θ, the probability of the event occurring in a single trial.

Classical approach to this problem.

PRIOR to observing the data, the probability of observing x was:

$$f(x|\theta) = \mathbb{P}(X = x \mid \theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}$$

$$x \in \{0, 1, \dots, n\}$$

This is a function of (the future) x for known θ.

If we know x, we could treat this as a function of θ, $L(\theta)$, the Likelihood function.

Suppose we are interested in a point estimate of θ

One strategy is to use the value of θ which maximises the likelihood. The rule is $\frac{x}{n}$ with corresponding estimator $\frac{X}{n} = T(X)$

Properties of the estimator (bias, consistency, ...) depend upon the sampling distribution of $T(X)$ given θ.

i.e. Distributions derived from $f(x \mid \theta)$

random fixed

[The distribution of $X|\theta$ e.g. $X|\theta \sim Bin(n, \theta)$]

* Such an approach can lead to nonsensical answers. *

EXAMPLE.

Suppose in the Bernoulli example, we wish to estimate θ^2.

An intuitive estimator (and also the mle) is $\left(\frac{X}{n}\right)^2$

This is biased.

$$\begin{aligned}
\mathbb{E}(X^2 \mid \theta) &= \mathrm{Var}(X \mid \theta) + \mathbb{E}^2(X \mid \theta) \\
&= n\theta(1-\theta) + n^2\theta^2 \\
&= n\theta + n(n-1)\theta^2 \\
&= \mathbb{E}(X \mid \theta) + n(n-1)\theta^2
\end{aligned}$$

i.e. $\theta^2 = \dfrac{\mathbb{E}(X^2 \mid \theta) - \mathbb{E}(X \mid \theta)}{n(n-1)}$, for $n > 1$

Thus, for $n > 1$, $\dfrac{X^2 - X}{n(n-1)} = \dfrac{X(X-1)}{n(n-1)}$ is an unbiased estimator of θ^2.

we see a success, so $\theta^2 \neq 0$

Suppose we observe $x = 1$. Our unbiased estimate of θ^2 is 0.

- We estimate a chance as zero even though the event has occurred.

Now let us consider two types of Bernoulli trials.

1. Toss a coin n times and observe x heads, parameter θ_c.

2. Toss a drawing pin n times and observe x pin-ups.

Parameter θ_p.

The roles for θ_c and θ_p are identical and share the same properties.

Classical statistics treats these two situations identically.

* Is this sensible? *